Research Statement, Dan Fabris

Our laboratory is dedicated to the **investigation of the structure-function relationships of protein-nucleic acid complexes involved in the lifecycle of viruses** responsible for infectious diseases, using mass spectrometry (MS) as the main instrumental platform. The greatest discoveries in biology and medicine have always been the fruit of significant technological advances and, vice versa, the greatest technological breakthroughs have always been prompted by the need to raise to the challenge of the biomedical problems of our times. The two go hand in hand in an indissoluble bond. It is this belief that permeates our research.

An example is provided by our **investigation of human immunodeficiency virus type 1 (HIV-1)**, the etiologic agent of AIDS. In this retrovirus, the 5'-untranslated region (5'-UTR) governs key replication steps in which viral RNA acts either as genome intended for packaging, or as mRNA meant for protein expression. The functions associated with these distinct spheres of activities are controlled by signals included in discrete elements of secondary structure, which define the overall 3D structure by establishing mutual tertiary interactions. *In vitro* experiments have provided evidence that 5'-UTR may assume different folds characterized by significant rearrangement of secondary structure, tertiary interactions, and overall architecture. Transitions between conformers are facilitated *in vitro* by the chaperone activity of the nucleocapsid (NC) domain of the Gag polyprotein, which is known for promoting remodeling of nucleic acid structure. These observations have prompted the hypothesis that NC/Gag may represent the actuator of a riboswitch mechanism that coordinates the multifaceted activities of 5'-UTR by exposing different signals at different points of the virus lifecycle. This hypothesis presupposes a process of specific recognition between unique structures present in the alternative conformers and cognate viral/host factors involved in the various activities, but the determinants of such interactions are still not understood. A variety of approaches including phylogenetic analysis, site-directed mutagenesis, knockdown experiments, and other biological assays have demonstrated different levels of functional correlation between 5'-UTR signals and cognate proteins. Explaining these functional relationships, however, has been hampered by the absence of actual structural data for pertinent protein-RNA assemblies implicated in such functions, which is largely due to the challenges of investigating these substrates directly in their natural environment. Despite extensive efforts, the dearth of *in vivo* structural information has limited our understanding of the determinants driving the recognition of specific viral/host factors, thus hampering their utilization as possible targets for rational drug design.

We propose that the missing information could be obtained by **combining chemical probing approaches with mass spectrometry detection (MS)**. Treating target substrates with these types of reagents resembles exposing photographic film to light. The process leaves a permanent "impression" of the substrate structure in the form of a modification pattern specific to the 3D fold, which can be subsequently "developed" by any suitable means that do not necessarily have to preserve the original fold. Taking advantage of this favorable characteristic, we are exploring the application of structural probing techniques to viral particles and infected cells. Once virions/cells have been exposed to the probe, they can be lysed to initiate the "development" process through experimental procedures that may employ

denaturing conditions. For example, the target ribonucleoprotein can be extracted by affinity capture using beads that are coated with suitable antisense oligonucleotides. The captured RNA can undergo cleavage by specific deoxyribozymes to isolate the region of interest. The probed products can be finally characterized by mass mapping and sequencing technologies. The MS platform is directly applicable to virtually all biomolecules and their mixtures, affording numerous advantages over techniques that require chromophores or probe-specific cleavage. The fact that the elemental composition of intact molecular ions and corresponding gas-phase fragments are very characteristic for each class of biopolymers and modifiers, alike, translates into the ability of achieving their full and unambiguous characterization from their unique mass signatures. We have been exploring MS approaches to facilitate the analysis of RNA-RNA and protein-RNA conjugates produced by bifunctional crosslinkers, which offer information about the spatial relationships between bound components. We are devising approaches capable of providing a direct view of binding interfaces, which are based on gas-phase processes rather than on probe modifications. We are also engaged in the development of software tools that support the interpretation of these types of data and enable the effective utilization of the corresponding spatial constraints in molecular modeling operations. Recent advances by computational techniques in structural biology have helped close the resolution gap by providing atomic-level details that are typically beyond the reach of chemical probes, thus dramatically increasing the value and reach of this type of approach for structural determination. We are taking advantage of these advances to pursue the elucidation of 5'-UTR and its functional ribonucleoproteins directly in their natural environments, which could provide the keys to understanding the mechanism by which this system performs its critical biological activities.

At the same time, we have been **exploring critical components of HIV and analogous retroviral systems as possible targets for the development of small ligand inhibitors**. The possibility of detecting intact non-covalent interactions by electrospray ionization (ESI) mass spectrometry has presented us with the opportunity to investigate the binding properties of chemicals that lack the chromophores necessary for their spectroscopic detection. Taking advantage of this favorable feature, we have developed technologies for determining the stoichiometry and binding affinity of ligands toward RNA substrates. We found new ways for determining the concentration of ligand that induces 50% inhibition of a target protein-RNA complex (i.e., IC50). The high resolution afforded by Fourier transform ion cyclotron resonance (FTICR) mass spectrometry has allowed us to investigate multiple ligands simultaneously in competitive binding experiments that can immediately identify the tightest binder in the pool and rank the relative binding affinities of the remaining ones. In proof-of-principle experiments, we created a random combinatorial library containing ~45,000 hepta-peptides, which was fractionated in three pools based on solubility in water, methanol, and 1:1 water:methanol. Each pool was then screened against a mixture of four selected structures of HIV-1 5'-UTR, which revealed the presence of ~50 species with superior binding affinity toward the different targets. These operations required ~30 min and resulted in reducing the number of potential drug candidates to the point where expensive *in vivo* screening becomes economically viable. We are now pursuing the implementation of front-end robotics and microfluidics systems to enable unattended high-throughput operations, which will allow us to explore a wider swath of chemical space for promising compounds. We believe that only

by developing this and other novel technologies will it be possible to accelerate the pace of drug discovery to keep up with emergence of new infectious diseases and the selection of drug-resistant strains.